Travis Gagie Dalhousie University

> DSB 2021 12/02/2021\*

<sup>\*</sup>Palindrome! :o)

### Travis Gagie

### Wheeler graphs

definition

example

РЫО

r indovina

, macxin

seeking TA

#### combination

.

mnomonia

------

## tunnel

definition

examples

mnomonics

hand-wavir

doubt

specular

hope?

questions

# Wheeler graphs

mnemon

seeking

r-inde

seeking TA

combinatio

mnemonic

motivation

tunne

definition

chimeric matcl

hand-wavin

doubt

hone?

uestions?

# definition

A Wheeler graph is a directed edge-labelled graph whose vertices can be ordered such that those with in-degree 0 precede those with positive in-degree and, for any pair of edges e=(u,v) and e'=(u',v') labelled a and a' respectively,

- if  $a \prec a'$  then v < v',
- if a = a' and u < u' then  $v \le v'$ .

### Wheeler graphs

deficial co

example

-------

FBID

mnemonic

seeking TA

#### r-indexing

#### Combinatio

result

mnemonic

motivation

#### tunneis

definition

examples

chimeric mater

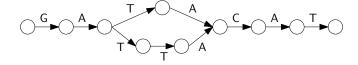
## hand-wavi

GOUDE

specula

questions

# example



### Wheeler graphs

deficial co

example

.

.

seeking TA

#### r-indexing

7-IIIGEXIII

rocult

Secring 171

#### combinatio

.

.

motivation

#### tunnels

1.0.30

examples

chimeric match

mnemonics

### nand-wavi

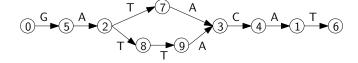
doubt

specula

......

#### questions

# example



Wheeler graphs

definition

exampl

PhID

mnemoni

seekiiig 17

rocult

seeking TA

combinatio

mnemonic

motivation

dofinition

examples chimeric match

mnemonics

nand-wavir

doubt

hope?

uestions

# result

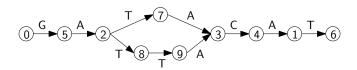
# Theorem

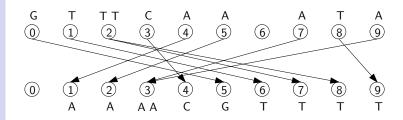
If G is a Wheeler graph and we store a rank data structure over its edge-labels, ordered by their origins, and partial sum data structures over its vertices' in- and out-degrees then later, given a pattern P, in  $O(|P|\log\log|G|)$  time we can find the interval in the Wheeler order containing vertices reachable by paths labelled P.

РЫД

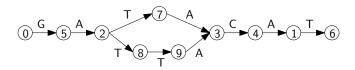
seeking TA

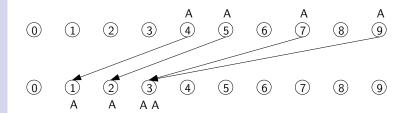
seeking TA



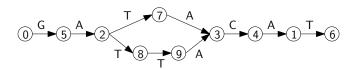


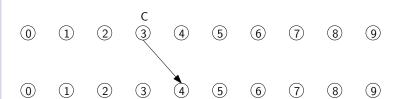
РЫД



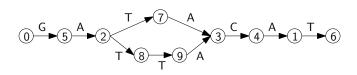


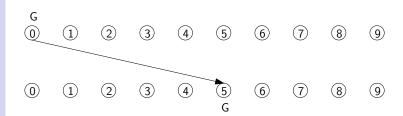
РЫД





РЫД





## Wheeler graphs

definitio

result

РЫО

sooking T

r-indexing

·

seeking TA

#### combinatio

\_\_\_\_\_

motivation

definition

examples

chimeric match

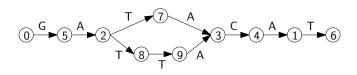
mnemonics

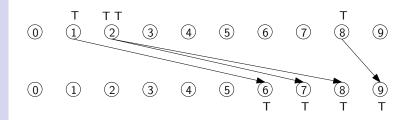
doubt

doubt

hope?

uestions





# Travis Gagie

Wheeler graphs

definition example

PbID

mnemonic

seeking T

r-indexin

seeking TA

combinatio

result mnemonics

motivatio

definition examples

chimeric match mnemonics

doubt

hope?

questions

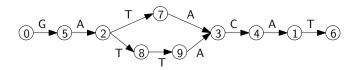
# mnemonic

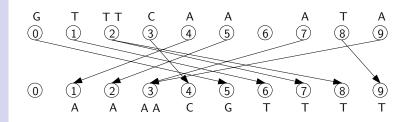


# Travis Gagie

seeking TA

seeking TA





## Travis Gagie

### Wheeler graphs

definitio

result

\_\_\_\_

seeking TA

#### seeking i

r-indexin

result

seeking TA

#### combinatio

.

motivation

#### tunne

definition

examples

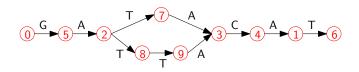
chimeric mate

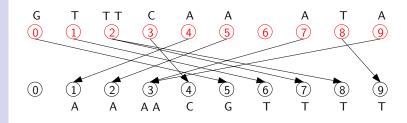
hand-wav

goupt

hone?

uestions





# Travis Gagie

### Wheeler graphs

definitio

result

\_\_\_\_

seeking TA

7-IIIdexIII

result

seeking TA

#### combinatio

-----

mnemonio

motivatio

#### tunne

definition

examples

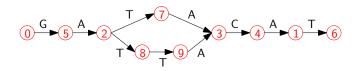
chimeric mate

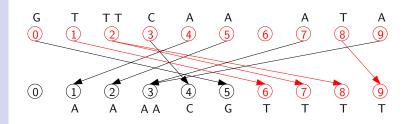
# hand-way

speculatio

hope?

uestions?





## Travis Gagie

### Wheeler graphs

definitio

result

seeking TA

#### r indovin

/ IIIGCXIII

result

seeking TA

#### combinatio

resure

motivation

### tunne

definition

examples

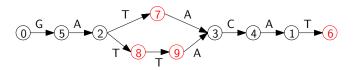
chimeric mate

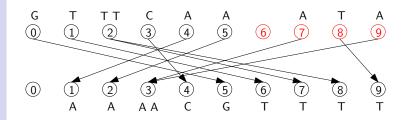
hand-wav

doubt

L---2

uestions





## Travis Gagie

### Wheeler graphs

definitio

exam

PЫD

mnemor

#### seeking TA

r-indexin

rocult

seeking TA

#### combinatio

mnemoni

motivation

#### tunne

definition

-----

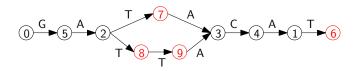
chimeric matc

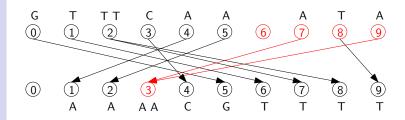
# doubt

speculatio

hone?

uestions





## Travis Gagie

### Wheeler graphs

definition

result

mnemon

#### seeking TA

r-indexin

result

seeking TA

#### combinatio

-----

mnemonio

motivatio

## tunne

definition

examples

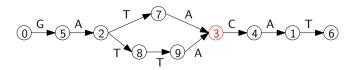
chimeric matc

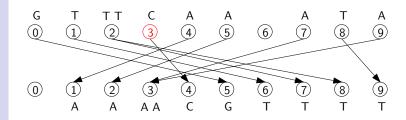
# doubt

goupt

hope?

uestions?





# Travis Gagie

### Wheeler graphs

definition

example

DLID

mnemonic

seeking TA

## r-indexing

......

seeking TA

#### combination

result

mnemonics

motivation

#### tunnels

definition

examples

ciminene in

hand .....

doubt

speculat

hope?

questions

# *r*-indexing

result

# result

# Theorem

Let S be a string and r be the number of runs in a BWT of S. We can store S in O(r) space such that later, given a pattern P, in  $O(|P| \log \log |S|)$  time we can find the interval in the BWT containing characters preceding occurrences of P in S, and then report the locations of those occurrences in  $O(\log \log |S|)$  time per occurrence.

## Wheeler graphs

. . . . .

exampl

result

mnemon

seeking T

*r*-indexi

result

44.00

COMBINACIO

mnemonic

motivation

tunnel

examples

chimeric match

hand-wavin

dente man

hono?

uestions

# result

# Theorem

Let S be a string and r be the number of runs in a BWT of S. We can store S in O(r) space such that later, given a pattern P, in  $O(|P|\log\log|S|)$  time we can find the interval in the BWT containing characters preceding occurrences of P in S, and then report the locations of those occurrences in  $O(\log\log|S|)$  time per occurrence.

Nishimoto & Tabei: total time is O(|P| + occ) when the alphabet size is polylog(|S|).

### Wheeler graphs

definitio

exampl

DUID

mnemonio

seeking Ta

r-indexi

#### result seeking TA

#### combinatio

COMBINACIO

mnemonic

motivation

#### tunnels

definition

examples

cnimeric match

## hand-wavi

doubt

hone?

uestions?

$$0 \xrightarrow{\mathsf{T}} \Theta \xrightarrow{\mathsf{A}} 0 \xrightarrow{\mathsf{G}} \Theta \xrightarrow{\mathsf{C}} \bullet 0 \xrightarrow{\mathsf{C}} \bullet 0 \xrightarrow{\mathsf{A}} \bullet 0 \xrightarrow{\mathsf{T}} \bullet 0 \xrightarrow{\mathsf{T}} \bullet 0 \xrightarrow{\mathsf{C}} \bullet 0 \xrightarrow{\mathsf{T}} \bullet 0 \xrightarrow{\mathsf{C}} \bullet 0 \xrightarrow{\mathsf{C}} \bullet 0 \xrightarrow{\mathsf{T}} \bullet 0 \xrightarrow{\mathsf{C}} \bullet 0 \xrightarrow{\mathsf{C}}$$

### Wheeler graphs

definitio

example

result

mnomonio

seeking TA

*r*-indexi

#### result seeking TA

#### combinatio

COMBINACIO

mnemonic

motivation

#### tunnels

definition

examples

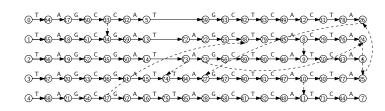
chimeric mater

## hand-wavi

doubt

speculati

uestions



### Wheeler graphs

definitio

exampl

result

.

seeking TA

r-indexi

result seeking TA

#### COMBINALIC

mnemonic

motivation

### tunnels

definition

examples

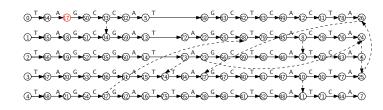
chimeric match

## hand-wavi

-----

hone?

uestions



# Travis Gagie

### Wheeler graphs

4-6-141---

exampl

rocult

seeking TA

seeking 17

result

seeking TA

#### combination

#### COMBINGUIC

result

------

motivation

#### . . . . . .

-----

chimeric matcl

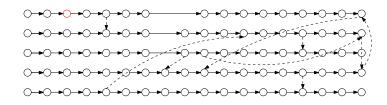
mnemonics

hand-wavi

doubt

. .

uestions



# Travis Gagie

### Wheeler graphs

definition

exampl

result

milemonic

seeking TA

result

seeking TA

Section 171

#### COMBINALIC

result

mnemoni

motivation

#### tunneis

definition

chimeric mate

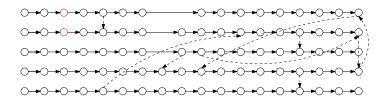
mnemonics

## hand-wavi

-----

. .

uestions



# Travis Gagie

### Wheeler graphs

definition

exampl

result

...

seeking TA

result

seeking TA

seeking TA

#### combinatio

result

mnemoni

motivation

#### tunnels

definition

chimeric matc

mnemonics

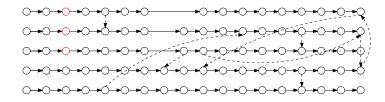
nand-wavi

doubt

speculatio

questions

uestions



# Travis Gagie

### Wheeler graphs

definition

exampl

rocult

seeking TA

seeking I/

rosult

seeking TA

#### combinati

#### COMBINGUIC

mnemonics

motivation

#### tunnels

definition

-----

chimeric mater

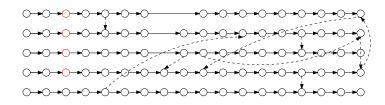
hand-wavi

doubt

speculatio

.....

uestions



# Travis Gagie

### Wheeler graphs

definition

exampl

result

.

seeking TA

Jeening 11

result

seeking TA

#### combinatio

#### COMBINACIO

mnemonic

motivation

#### tunnels

definition

chimoric matel

mnemonics

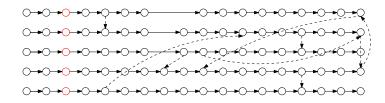
## nand-wavii

doubt

speculatio

. .

uestions



# Travis Gagie

### Wheeler graphs

1.0.50

example

result

.

seeking TA

result

seeking TA

seeking 17

#### COMBINALIC

result

mnemoni

motivation

#### . . . . . .

definition

chimeric matc

mnemonics

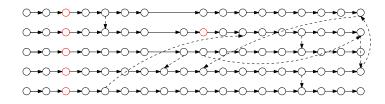
### nand-wavii

doubt

speculati

.....

uestions



# Travis Gagie

### Wheeler graphs

1.0.50

exampl

rocult

. ...

seeking TA

Seeking 17

result

seeking TA

# seeking TA

#### combinatio

result

mnemoni

motivation

#### turmer

definition

-----

mnemonics

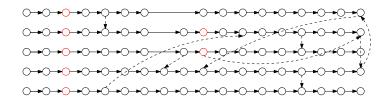
hand-wavi

doubt

speculatio

hope?

uestions



# Travis Gagie

### Wheeler graphs

definition

exampl

result

seeking TA

Jeening 11

result

seeking TA

#### COMBINALIC

result

mnemonio

motivation

#### 4-6-141--

examples

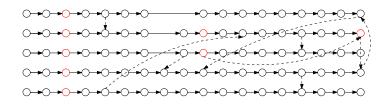
chimeric match

hand-wavi

doubt

speculatio

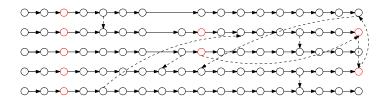
uestions



# Travis Gagie

seeking TA

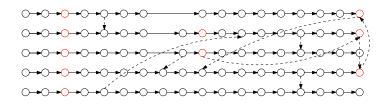
seeking TA



# Travis Gagie

seeking TA

seeking TA



# Travis Gagie

### Wheeler graphs

dofinition

example

result

\_\_\_\_\_

seeking TA

Jeening 17

-----

seeking TA

#### combinatio

#### COMBINGUIC

mnemonic

motivation

#### tunnel

definition

-1-1------

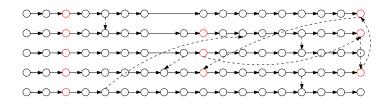
mnemonics

# hand-wavi

doubt

speculatio

uestions



# Travis Gagie

### Wheeler graphs

dofinition

exampl

result

seeking TA

result

seeking TA

#### combinatio

#### COMBINGLIC

mnemonic

motivation

#### tunnels

definition

chimoric matel

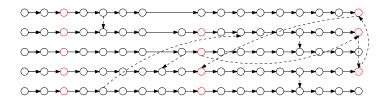
chimeric match

# hand-wavii

-----

hone?

uestions



#### Wheeler graphs

definitio

exampl

DUID

mnemonio

seeking TA

r-index

#### result seeking TA

#### combinatio

COMBINACIO

mnemonic

motivation

#### tunnels

definition

examples

chimeric match

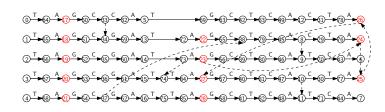
#### hand-wavi

1 ...

. .

uestions

# seeking TA



### Travis Gagie

#### Wheeler graphs

definition

example

PHD

mnemonic

seeking TA

#### r-indexing

rocult

seeking TA

#### combination

resuit

-----

#### tunnel

4-6-141--

examples

ciiiiiciic iii

mnemonics

### IIaiiu-wavii

hono?

questions

# combination

### Wheeler graphs

definition

examp

mnema

seeking T

r-index

seeking T

combinatio

result

mnemonic

tunnole

definition

chimeric match

hand-wavin

Janea Warn

specula

uestions?

### result

### Theorem

Let G be a Wheeler graph and r be the number of runs in a BWT of G, and suppose G can be decomposed into v edge-disjoint directed paths whose internal vertices each have in- and out-degree exactly 1. We can store G in O(r+v) space such that later, given a pattern P, in  $O(|P|\log\log|G|)$  time we can count the vertices of G reachable by directed paths labelled P, and then report those vertices in  $O(\log\log|G|)$  time per vertex.

### Wheeler graphs

definition

examp

PbID

mnemonic

r-index

result seeking TA

combinati

result

mnemonic motivation

tunnel

examples chimeric match

mnemonics

Hallu-waviii

specul

uestions

### Theorem

Let G be a Wheeler graph and r be the number of runs in a BWT of G, and suppose G can be decomposed into v edge-disjoint directed paths whose internal vertices each have in- and out-degree exactly I. We can store G in O(r+v) space such that later, given a pattern P, in  $O(|P|\log\log|G|)$  time we can count the vertices of G reachable by directed paths labelled P, and then report those vertices in  $O(\log\log|G|)$  time per vertex.

Can we use Nishimoto & Tabei's technique to get O(|P| + occ) time when the alphabet size is polylog(|G|)?

#### Wheeler graphs

definition

exampl

РЫД

mnemonic

seeking TA

#### r-indexing

result

seeking TA

#### combinatio

#### .

mnemonics

. . . . . .

motivation

#### tunnels

definition

examples

cnimeric matci

hand-wavi

doubt

speculatio

.....

luestions?

## mnemonics

- 1  $\upsilon$  ("upsilon") is the number of unitigs
- 2 Travis says \ups ("oops") a lot

#### Travis Gagie

Wheeler graphs

definition

examp

result

mnemoni

seeking T

r-inde

seeking T

combinatio

COMBINACIO

mnemor

motivation?

motivatio

J.C.:ki-

examples

chimeric match

hand-wavin

doubt

hone?

uestions

## motivation?

We needed a special case of this theorem in a recent paper Garance will present, about long and stringy trees.

It's easier to prove the general theorem than to explain why we'd want it. Wheeler graphs are cool and r-indexes are cool, so r-indexing Wheeler graphs is really cool?

Jouni says he's already implemented something like this, so it must be useful for something. . .

Nicola has a more sophisticated result in a SODA '21 paper (but I'm not sure it's practical).

### Travis Gagie

#### Wheeler granhs

definition

example

01.10

mnomonio

seeking TA

#### r-indexing

result

seeking TA

#### and the section of

......

mnemonic

motivation

#### tunnels

definition

examples chimeric match

mnemonics

hand-wavir

. . .

hono?

questions

# tunnels

#### Travis Gagie

#### definition

### definition

Uwe Baier introduced tunnelling to reduce the number of times we write large run-lengths in the RLBWT.

We can put paths in a Wheeler graph together in a tunnel if

- corresponding edge labels match
- corresponding vertices' ranks are incremented in the Wheeler order

We stop merging "one edge too early".

### Travis Gagie

#### Wheeler graphs

definition

exampl

result

.

seeking TA

...

/ IIIGC/

seeking TA

#### Combinatio

result

mnemoni

motivatio

#### tunne

examples

chimeric mate

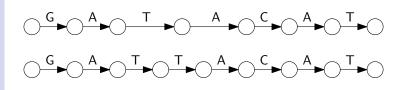
mnemonics

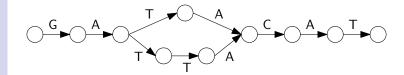
dente man

------

hone?

uestions





#### Wheeler graphs

definitio

exampl

DEID

mnomoni

seeking T.

r-indexi

result

seeking TA

#### combinatio

rocult

mnemonics

motivation

#### tunnels

examples

chimeric mate

hand-wavi

doubt

speculati

nuections

uestions?

$$0 \xrightarrow{\mathsf{T}} \bullet 0 \xrightarrow{\mathsf{A}} 0 \xrightarrow{\mathsf{G}} \bullet 0 \xrightarrow{\mathsf{C}} \bullet 0 \xrightarrow{\mathsf{G}} \bullet 0 \xrightarrow{\mathsf{$$

#### Wheeler graphs

definitio

example

PHD

mnemonic

seeking TA

r-indexi

result

seeking TA

#### combinatio

......

mnemonics

motivation

#### tunnel

examples

chimeric match

hand-wavi

doubt

. .

uestions?

$$0 \xrightarrow{\mathsf{T}} + 0 \xrightarrow{\mathsf{A}} 0 \xrightarrow{\mathsf{G}} + 0 \xrightarrow{\mathsf{C}} + 0 \xrightarrow{\mathsf{$$

#### Wheeler graphs

definitio

example

resure

mnomonic

seeking TA

r indov

result

seeking TA

#### combinatio

.

mnemonio

motivation

#### tunnel

examples

chimeric match

chimene mater

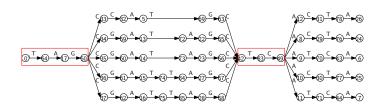
#### hand-wavi

doubt

specula

auestions

#### uestions



#### Wheeler graphs

dofinitio

example

PHID

mnemonic

seeking TA

Julian I

result

seeking TA

combinatio

\_\_\_\_\_

.

.. ..

motivation

#### turmer

#### examples

chimeric matcl

mnomonics

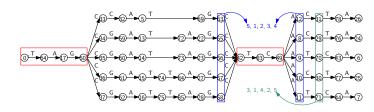
#### hand-wavi

doubt

specula

questions

#### uestions?



#### Travis Gagie

#### Wheeler graphs

definition

examp

PbID

mnemoni

r\_indev

result seeking T/

#### combinatio

result

motivation

tunnels

definitio

chimeric matches

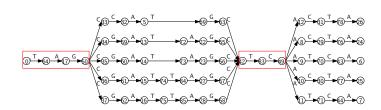
hand-wavin

doubt

hone?

uestions

## chimeric matches



### **Theorem**

We can index a tunnelled Wheeler graph so as to prevent false-positive, chimeric matches.

(In our example, if the first variation is CAT, the second variation *isn't* TC.)

#### Travis Gagie

#### Wheeler graphs

definition

result

seeking TA

*r*-indexi

result seeking TA

#### combinatio

result

mnemonics

motivation

#### tunnel

definition examples

cimilaria in

mnemonics

doubt

speculatio

hope?

uestions

## mnemonics



2 tunnels are FIFO

#### Travis Gagie

#### Wheeler graphs

definition

example

PHD

mnemonic

seeking TA

#### r-indexing

------

seeking TA

#### combination

......

mnomonio

motivation

#### tunnel

definition

examples

\_\_\_\_

mnemonics

### hand-waving

doubt

. . . .

hope?

uestions

# hand-waving

#### Wheeler graphs

definitio

examp

РЫД

mnemonic

seeking TA

result

seeking T/

#### combination

mnemon

motivation

#### tunnels

definition

chimeric mat

mnemonics

#### hand-wa

doubt

. .

uestions

## doubt

Would we want to r-index a tunnelled Wheeler graph?

Doesn't tunnelling the graph collapse the runs in the BWT?

Wouldn't a standard sampling be better then?

Travis Gagie

# speculation

# Suppose

• the conserved regions between variation sites are on the order of 1K bp

- the probability of a bp mutating is on order of 1 in 100M<sup>†</sup>
- we're indexing 1M genomes.

We'd like to tunnel the conserved regions, but how do we deal with the mutations?

speculation

†I've read this is the mutation rate in humans, but that seems to be the probability a bp is mutated between generations, not the probability of picking a bp and a person at random and having that bp mutated in that person. Any idea?

### Travis Gagie

#### Wheeler graphs

definition

result

mnemonic

seeking TA

r-indexing

seeking TA

#### combination

result

motivation

## tunne

examples

mnemonics

doubt

speculation

-

questions!

# speculation



(not recommended!)

#### Wheeler graphs

definition

result

mnemor

seeking T

r-indexi

seeking TA

#### combinatio

result

mnemonic

tunnels

definition

chimeric match

hand-wavin

speculation

...

uestions

# hope?

The "conserved" regions of about 1K bp are each broken up by about (1M  $\times$  1K/100M) = 10 mutations, in expectation.

If we merge the paths in each conserved region into about 10 tunnels, however, then we *should* get a significantly smaller, interesting Wheeler graph whose BWT is still highly run-length compressible?

#### Travis Gagie

#### Wheeler graphs

definition

rosult

PHD

mnemonic

seeking TA

r-indexing

result

seeking TA

#### combinati

mnomonic

motivation

tunners

definition

chimeric mate

\_\_\_\_

hand-wavi

doubt

speculatio

hope?

questions?

# questions?

Happy Year of the Ox!

Un saludo pa' Moni!